Expert Systems with Applications 31 (2006) 264–269

Expert Systems with Applications

www.elsevier.com/locate/eswa

A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia

Kemal Polat *, Seral Şahan, Salih Güneş

Department of Electrical & Electronics Engineering, Engineering-Architecture Faculty, Selcuk University, Konya 42075, Turkey

Abstract

Changes in the normal rhythm of a human heart may result in different cardiac arrhythmias, which may be immediately fatal or cause irreparable damage to the heart sustained over long periods of time. The ability to automatically identify arrhythmias from ECG recordings is important for clinical diagnosis and treatment. Artificial immune systems (AISs) is a new but effective branch of artificial intelligence. Among the systems proposed in this field so far, artificial immune recognition system (AIRS), which was proposed by A. Watkins, has showed an effective and intriguing performance on the problems it was applied. Previously, AIRS was applied a range of problems including machine-learning benchmark problems and medical classification problems like breast cancer, diabets, liver disorders classification problems. The conducted medical classification task was performed for ECG arrhythmia data taken from UCI repository of machine-learning. Firsly, ECG dataset is normalized in the range of [0,1] and is weighted with fuzzy weighted pre-processing. Then, weighted input values obtained from fuzzy weighted pre-processing is classified by using AIRS classifier system. In this study, fuzzy weighted pre-processing, which can be improved by ours, is a new method and firstly, it is applied to ECG dataset. Classifier system consists of three stages: 50–50% of traing-test dataset, 70–30% of traing-test dataset and 80–20% of traing-test dataset, subsequently, the obtained classification accuries: 78.79, 75.00 and 80.77%.

Keywords: ECG arrhythmia; Artificial immune system; AIRS; Fuzzy weighted pre-processing

1. Introduction

One of the central problems of the information age is dealing with the enormous amount of raw information that is available. More and more data is being collected and stored in databases or spreadsheets. As the volume increases, the gap between generating and collecting the data and actually being able to understand it is widening. In order to bridge this knowledge gap, a variety of techniques known as data mining or knowledge discovery is being developed. Knowledge discovery can be defined as the extraction of implicit, previously unknown, and potentially useful information from real world data, and communicating the discovered knowledge to people in an understandable way (Delen, Walker, & Kadam, 2004; Goodman, Boggess, & Watkins, 2003; Soman & Bobbie, 2005).

The motivation behind the research reported in this paper is the results obtained from extensions of an ongoing research effort. The research reported in Perelson and Oster (1979) and Pazzani and Kibler (1992) is on developing a non-invasive ECG hardware and embedded software for capturing, analysing, diagnosing, and recommending remedies for home-care patients with heart conditions. In the effort, we focused on the (hardware) acquisition and (software) analysis of ECG signals for early diagnosis of Tachycardia heart disease. The work reported here builds on the initial work by, first, using machine-learning techniques to study and understand the accurate prediction of arrhythmic diseases and suggestive remedies based on the classification schemes or models (Pazzani & Kibler, 1992; Perelson & Oster, 1979).

The paper is organized as follows: Section 2 introduces fuzzy weighted pre-processing. Section 3 was introduced from natural and artificial immune systems. In Section 4, the performance analysis of AIRS was presented for conducted application. Consequently in Section 5, the results were given

* Corresponding author. Tel.: +90 332 2232098; fax: +90 332 2410635.
E-mail addresses: kpolat@selcuk.edu.tr (K. Polat), seral@selcuk.edu.tr (S. Şahan), sgunes@selcuk.edu.tr (S. Güneş).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.09.019

and evaluated in addition to emphasizing the possible studies that can be done related with this study.

2. Background

2.1. ECG arrhythmia classification problem

The dataset used in this study was obtained from the archives of machine-learning datasets at the University of California, Irvine (Şahan, Kodaz, Güneş, & Polat, 2004). The datasets are grouped into three broad classes to facilitate their use in experimentally determining the presence or absence of arrhythmia, and for identifying the type of arrhythmia. In the set, Class 01 refers to 'normal' ECG. Classes 02–15 refer to different classes of arrhythmia and Class 16 refers to the rest of unclassified data. The arrhythmia dataset has 279 attributes, 206 of which are linear valued and the rest are nominal. There are 452 instances, and as indicated above, 16 classes. There are missing values in the dataset. In such cases, probabilistic values were assigned according to the distribution of the known values for the attributes.

2.2. Previous research

There has been much work in the field of classification and most work has been based on neural networks, Markov chain models and support vector machines (SVMs). The datasets used to train these methods are often small. In Pazzani and Kibler (1992), direct-kernel methods and support vector machines are used for pattern recognition in magnetocardiography. In Fayyad, Piatetsky-Shapiro, Smyth and Uthurusamy (1996), self-organizing maps (SOMs) are used for analysis of ECG signals. The SOMs helps discover a structure in a set of ECG patterns and visualize a topology of the data. In Michie (1991), machine-learning methods like artificial neural networks (ANNs) and logically weighted regression (LWR) methods are used for automated morphological galaxy classification. The focus of the investigation described in this paper is to evaluate three standard machine-learning algorithms applied to classify cardiac arrhythmias. All related previous research cited in this paper use classes, features, and machine-learning methods and related software, which we used. Therefore, our comparisons are in the context of the predictability, accuracy, and ease of learning of these algorithms. The former two capabilities are significant in diagnosing and treating ECG abnormalities while the latter facilitates the practical use of our ECG diagnostic device.

2.3. Natural and artificial immune systems

The natural immune system is a distributed novel-pattern detection system with several functional components positioned in strategic locations throughout the body. Immune system regulates defence mechanism of the body by means of innate and adaptive immune responses. Between these, adaptive immune response is much more important for us because it contains metaphors like recognition, memory acquisition, diversity, self-regulation, etc. The main architects of adaptive immune response are lymphocytes, which is divide into two classes as T and B lymphocytes (cells), each having its own function. Especially B-cells have a great importance because of their secreted antibodies (Abs*)* that take very critical roles in adaptive immune response.

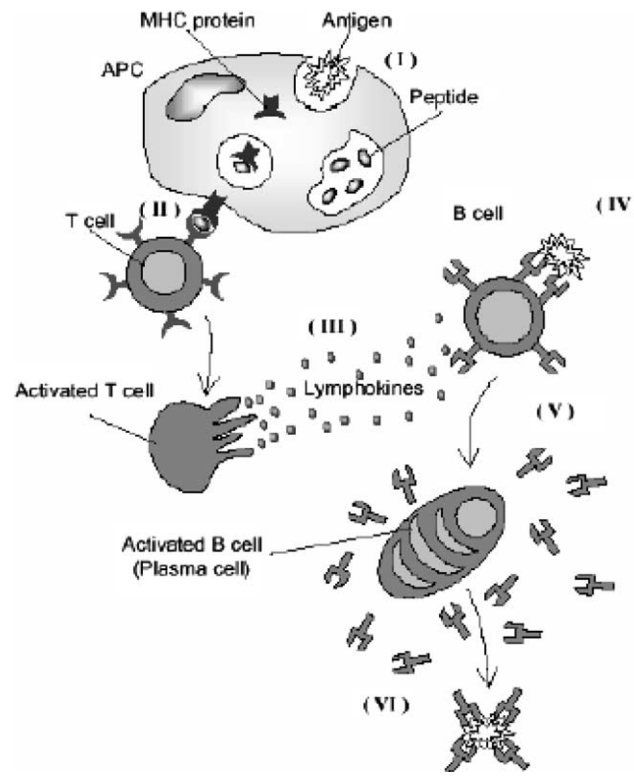

Fig. 1. Immune response (Perelson & Oster, 1979).

The simplified working procedure of our immune system is illustrated in Fig. 1. Specialized antigen presenting cells (APCs) called macrophages circulates through the body and if they encounter an antigen, they ingest and fragment them into antigenic peptides (I). The pieces of these peptides are displayed on the cell surface by major histocompatibility complex (MHC) molecules existing in the digesting APC. The presented MHC–peptide combination on the cell surface is recognised by the T-cells causing them to be activated (II). Activated T-cells secrete some chemicals as alert signals to other units in response to this recognition (III). B-cells, one of the units that take these signals from the T-cells become activated with the recognition of antigen by their antibodies occurred in the same time (IV). When activated, B-cells turn into plasma cells that secrete bound antibodies on their surfaces (V). Secreted antibodies bind the existing antigens and neutralize them signalling other components of immune system to destruct the antigen–antibody complex (VI) (de la Calleja & Fuentes, 2004). For detailed information about immune system refer to Abbas and Lichtman (2003); Watkins (2001).

Artificial immune systems emerged in the 1990s as a new computational research area. Artificial immune systems link several emerging computational fields inspired by biological behaviour such as artificial neural networks and artificial life.